

Governance Indicators:

Where Are We, Where Should We Be Going?

Daniel Kaufmann

Aart Kraay

The World Bank
World Bank Institute
Global Governance Group
and
Development Research Group
Macroeconomics and Growth Team



Abstract

Scholars, policymakers, aid donors, and aid recipients acknowledge the importance of good governance for development. This understanding has spurred an intense interest in more refined, nuanced, and policy-relevant indicators of governance. In this paper we review progress to date in the area of measuring governance, using a simple framework of analysis focusing on two key questions: (i) what do we measure? and, (ii) whose views do we rely on? For the former question, we distinguish between indicators measuring formal laws or rules 'on the books', and indicators that measure the practical application or outcomes of these rules 'on the ground', calling attention to the strengths and weaknesses of both types of indicators as well as the complementarities between them. For the latter question, we distinguish

between experts and survey respondents on whose views governance assessments are based, again highlighting their advantages, disadvantages, and complementarities. We also review the merits of aggregate as opposed to individual governance indicators. We conclude with some simple principles to guide the refinement of existing governance indicators and the development of future indicators. We emphasize the need to: transparently disclose and account for the margins of error in all indicators; draw from a diversity of indicators and exploit complementarities among them; submit all indicators to rigorous public and academic scrutiny; and, in light of the lessons of over a decade of existing indicators, to be realistic in the expectations of future indicators.

This paper—a joint product of the Global Governance Group, World Bank Institute, and the Macroeconomics and Growth Team, Development Research Group—is part of a larger effort in the Bank to study governance. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at dkaufmann@worldbank.org, akraay@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Governance Indicators: Where Are We, Where Should We Be Going?

Daniel Kaufmann
Aart Kraay

The World Bank

1818 H Street N.W., Washington, DC 20433, dkaufmann@worldbank.org, akraay@worldbank.org. We would like to thank Shanta Devarajan for encouraging us to write this survey for the World Bank Research Observer, three anonymous referees for their helpful comments, and Massimo Mastruzzi for assistance. The views expressed here are the authors' and do not necessarily reflect those of the World Bank, its Executive Directors, or the countries they represent.

***"Not everything that can be counted counts,
and not everything that counts can be counted"***

Albert Einstein

1. Introduction

Most scholars, policymakers, aid donors, and aid recipients recognize that good governance is a fundamental ingredient of sustained economic development. This growing understanding, which was initially informed by a very limited set of empirical measures of governance, has spurred an intense interest in developing more refined, nuanced, and policy-relevant indicators of governance. In this paper we review progress to date in the area of measuring governance, emphasizing empirical measures that are explicitly designed to be comparable across countries, and in most cases, over time as well. Our goal here is to provide a structure for thinking about the strengths and weaknesses of different types of governance indicators that can inform ongoing efforts to improve existing measures and develop new ones.¹

We begin in Section 2 by reviewing some of the alternative definitions of governance, as a necessary first step towards measurement. Although there are many broad definitions of governance in circulation, the degree of definitional disagreement can easily be overstated. Most definitions appropriately emphasize the importance of a capable state, accountable to its citizens and operating under the rule of law. Broad principles of governance along these lines are naturally not amenable to direct observation and thus to direct measurement: as the first part of the quote from Albert Einstein reminds us, "not everything that counts can be counted". However as we document below there are many different types of data that are informative of the extent to which these principles of governance are observed across countries. An important corollary is that any particular indicator of governance can usefully be interpreted as a noisy, or imperfect *proxy* for some unobserved broad dimension of governance. This interpretation emphasizes a recurrent theme throughout this review -- that there is

¹ We do not provide a great deal of detail on each of the many existing indicators of governance. All of the measures we discuss have been competently described by their producers, several have attracted their own written critiques and discussions, and there are already a number of existing surveys and user guides to the body of existing governance indicators. See for example Arndt and Oman (2006), Knack (2006), UNDP (2005), and Chapter 5 of World Bank (2006). Due to space constraints we also do not attempt to review the very important body of work focused on in-depth within-country diagnostic measures of governance that are not designed for cross-country replicability and comparisons.

measurement error in all governance indicators. This measurement error should be explicitly considered when using this kind of data to draw conclusions about cross-country differences or trends over time in governance.

We organize our discussion in Sections 3 and 4 around a simple taxonomy of existing governance indicators, summarized in Table 1. The first dimension of our taxonomy captures varying answers to the question "What do we measure?", that we take up in Section 3. We highlight the distinction between indicators that measure the existence of specific laws or rules 'on the books', and indicators that measure particular governance outcomes 'on the ground'. The former codifies details of the constitutional, legal or regulatory environment, the existence or absence of specific agencies such as anticorruption commissions or independent auditors, etc., that are intended to provide the key *de jure* foundations of governance. The latter are indicators that measure *de facto* governance outcomes that result from the of the application of these rules: for example, do firms find the regulatory environment cumbersome?, do households believe the police are corrupt?, etc.. An important message in this section concerns the shared limitations of indicators of both rules and outcomes: outcome-based indicators of governance can be difficult to link back to specific policy interventions, and conversely, the links from easy-to-measure *de jure* indicators of rules to governance outcomes of interest are in many cases not yet well-understood, and in some cases appear tenuous at best. The second part of the Einstein quote reminds us of the need for modesty in this respect: "not everything that can be counted counts".

The other dimension of our taxonomy corresponds to varying answers to the question "Whose views do we rely on?", that we take up in Section 4. We distinguish between indicators based on the views of various types of experts, and those survey-based indicators that capture the views of large samples of firms and individuals. In addition we identify a category of aggregate indicators that combine, organize, and summarize information from these different types of respondents. Section 5 of the paper is devoted to discussing the rationale for, and strengths and weaknesses of, such aggregate indicators.

The entries in Table 1 are a selection of existing governance indicators that we discuss throughout the paper. The table entries are not intended to be exhaustive of the

stock of existing governance indicators, but rather as leading examples of major indicators in this taxonomy.² A striking feature of efforts to measure governance to date is the preponderance of indicators focused on measuring various *de facto* governance outcomes, contrasting the relative few which measure *de jure* rules. Almost by necessity, the latter type of rules-based indicators of governance reflects the views or judgments of *experts* in the relevant areas. In contrast, the much larger body of *de facto* indicators captures the views both of experts as well as survey respondents of various types.

We conclude in Section 6 with a discussion of the way forward with measuring governance in a manner that can be useful to policymakers. We emphasize the importance of consumers and producers of governance indicators clearly recognizing and disclosing the pervasive measurement error in *all* types of governance indicators. We also note that to further a constructive discussion on governance indicators it is important to move away from oft-heard false dichotomies, such as ‘subjective’ vs. ‘objective’ indicators, or aggregate vs. disaggregated ones. As we discuss below virtually all measures of governance, for good reason, involve a degree of subjective judgment. And with respect to aggregation, different levels of aggregation are appropriate for different types of analysis, and in any case this is not an either-or distinction as most aggregate indicators can readily be unbundled into their constituent components.

We also emphasize the importance of both broad public scrutiny as well as more narrow and technical scholarly peer review of governance indicators. And finally, our overall conclusion is that while there has been considerable progress in the area of measuring governance over the past decade, the indicators that exist, and the ones that are likely to emerge in the near future, will remain imperfect. This in turn underscores the importance of relying on a diversity of the different types of indicators when monitoring governance and formulating policies to improve governance.

² For access to a fuller compilation of governance datasets, visit www.worldbank.org/wbi/governance/data

2. What Do We Mean By "Governance"?

The concept of governance is not new. Early discussions go back to at least 400 B.C. to the Arthashastra, a fascinating treatise on governance attributed to Kautilya, thought to be the chief minister to the King of India. In it, Kautilya presented key pillars of the 'art of governance', emphasizing justice, ethics, and anti-autocratic tendencies. He further detailed the duty of the king to protect the wealth of the State and its subjects; to enhance, maintain and also safeguard such wealth, as well as the interests of the subjects.

Despite the long provenance of the concept, there is as yet no strong consensus around a single definition of governance or institutional quality. In the spirit of this absence of consensus, throughout this paper we use interchangeably, even if somewhat imprecisely, the terms "governance", "institutions", and "institutional quality". Various authors and organizations have produced a wide array of definitions. Some are so broad that they cover almost anything, such as the definition of "rules, enforcement mechanisms, and organizations" offered by the World Bank's 2002 World Development Report "Building Institutions for Markets".³ Others like the one offered by Douglass North, are not only broad, but risk making the links from good governance to development almost tautological:

*"How do we account for poverty in the midst of plenty? We must create incentives for people to invest in more efficient technology, increase their skills, and organize efficient markets Such incentives are embodied in institutions"*⁴

As we discuss further below, some of the governance indicators we survey are similarly broad in that they capture a wide range of development outcomes as well. While we recognize that it is difficult to draw a bright line between governance and ultimate development outcomes of interest, we think it is useful at both the definitional and measurement stages to emphasize concepts of governance that are at least somewhat removed from development outcomes themselves. For example, an early and narrower definition of public sector governance proposed by the World Bank in 1992 is that:

*"Governance is the manner in which power is exercised in the management of a country's economic and social resources for development"*⁵

³ World Bank (2002), p. 6.

⁴ North (2000).

In the Bank's latest governance and anticorruption strategy, this definition has persisted almost unchanged, with governance defined as:

"...the manner in which public officials and institutions acquire and exercise the authority to shape public policy and provide public goods and services".⁶

In our own work on aggregate governance indicators that we discuss further below, we defined governance drawing on existing definitions as:

"...the traditions and institutions by which authority in a country is exercised. This includes the process by which governments are selected, monitored and replaced; the capacity of the government to effectively formulate and implement sound policies; and the respect of citizens and the state for the institutions that govern economic and social interactions among them."⁷

While the many existing definitions of governance cover a broad range of issues, one should not conclude that there is a total lack of definitional consensus in this area. Most definitions of governance agree on the importance of a capable state operating under the rule of law. Interestingly, comparing the last three definitions provided above, the one substantive difference has to do with the explicit degree of emphasis on the role of democratic accountability of governments to their citizens. And even these narrower definitions remain sufficiently broad that there is scope for a wide diversity of empirical measures of various dimensions of good governance.

The gravity of the issues dealt with in these various definitions of governance suggests that measurement in this area is important. While less so nowadays, in recent years there has however been considerable debate as to whether such broad notions of governance can in fact be usefully measured. Here we make a simple and fairly uncontroversial observation: there are many possible indicators that can shed light on various dimensions of governance. However, given the breadth of the concepts, and in many cases their inherent unobservability, no one indicator, or combination of indicators, can provide a completely reliable measure of any of these dimensions of governance. Rather, it is useful to think of the various specific indicators that we discuss below as all

⁵ World Bank (1992)

⁶ World Bank (2007), p. i, para. 3.

⁷ Kaufmann, Kraay, and Zoido-Lobaton (1999), p.1.

providing noisy or imperfect signals of fundamentally unobservable concepts of governance. This interpretation emphasizes the importance of taking into account as explicitly as possible the inevitable resulting measurement error in all indicators of governance when analyzing and interpreting any such measure. As we shall see below, however, the fact that such margins of error are finite and still allow for meaningful country comparisons both across space and time does suggest that governance measurement is both feasible and informative.

3. What Do We Measure: Governance Rules or Governance Outcomes?

In this section we discuss, in turn, rules-based indicators of governance, and outcome-based indicators of governance. To illustrate this distinction consider possible alternative measures of corruption. At the one extreme of rules-based indicators we can measure whether countries have legislation that prohibits corruption, or whether an anticorruption agency exists. But we can also measure whether in practice, the laws regarding corruption are enforced, or whether the anticorruption agency is undermined by political interference. And going one step further one can collect information on the views of firms, individuals, NGOs, or commercial risk rating agencies regarding the prevalence of corruption in the public sector.

Similarly for public sector accountability, we can observe rules regarding the presence of formal elections, financial disclosure requirements for public servants, and the like. But one can also assess the extent to which these rules operate in practice, and one can obtain information on the views of respondents as to the functioning of the institutions of democratic accountability. We first discuss these rules-based or *de jure* indicators of governance, and then turn to the outcome-based or *de facto* indicators. Clearly, at times there is no "bright line" dividing the two types, and so it is more useful to think of ordering different indicators along a continuum, with one end corresponding to rules and the other to ultimate governance outcomes of interest. Since both types of indicators have their strengths and weaknesses, we emphasize at the outset that all of these indicators should be thought of as imperfect, but complementary, proxies for the aspects of governance that they purport to measure.

Rules-Based Indicators of Governance

Several well-known examples of rules-based indicators of governance are noted in Table 1, including the Doing Business project of the World Bank, which reports detailed information on the legal and regulatory environment in a large set of countries; the Database of Political Institutions constructed by World Bank researchers, and also, the POLITY-IV database of the University of Maryland that both report detailed factual information on the features of countries' political systems; and the Global Integrity Index which provides detailed information on the legal framework governing public sector accountability and transparency in a sample of 41 mostly developing countries.

At first glance, one of the main virtues of indicators of rules is their clarity. It is straightforward to ascertain whether a country has a presidential or a parliamentary system of government, or whether a country has a legally-independent anticorruption commission. In principle it is also straightforward to document details of the legal and regulatory environment, such as how many distinct legal steps are required to register a business or to fire a worker. This clarity also implies that it is straightforward to measure progress on such indicators: Has an anticorruption commission been established? Have business entry regulations been streamlined? Has a legal requirement for disclosure of budget documents been passed? This clarity has made such indicators very appealing to aid donors interested in linking aid with performance indicators in recipient countries, and in monitoring progress on such indicators.

Set against these advantages are what we see as three main drawbacks. First, it is easy to overstate the clarity and objectivity of rules-based measures of governance. In practice there is a good deal of subjective judgment involved in codifying all but the most basic and obvious features of countries' constitutional, legal, and regulatory environments. After all, it is no accident that the views of lawyers -- on which many of these indicators are based -- are commonly referred to as "opinions". For example, in Kenya at the time of writing, a constitutional right to access to information may be undermined or offset entirely by an official secrecy act and by pending approval and implementation of the Freedom of Information Act, so that codifying even the legal right to access to information requires careful judgment as to the net effect of potentially conflicting laws. Of course, this drawback of ambiguity is hardly unique to rules-based

measures of governance: as we discuss below interpreting outcome-based indicators of governance can also involve significant ambiguities. However, for rules-based indicators in particular there has been less recognition of the extent to which they are also based on subjective judgment.

A second drawback of this type of indicator follows from the simple observation that the links from such indicators to outcomes of interest are complex, possibly subject to long lags, and often not well-understood. This complicates the interpretation of rules-based indicators. And of course, as we discuss below, symmetric difficulties arise in the interpretation of outcome-based indicators of governance, which can be difficult to link back to specific legal policy levers.

In the case of rules-based measures, some of the most basic features of countries' constitutional arrangements have little normative content on their own; instead such indicators are for the most part descriptive. For example, it makes little sense to presuppose that presidential (as opposed to parliamentary) systems, or majoritarian (as opposed to proportional) representation in voting arrangements, are intrinsically "good" or "bad" on their own. Rather the interest in such variables as indicators of governance rests on the case that they may matter for outcomes, often in complex ways. In an influential recent book, for example, Persson and Tabellini (2005) document how these features of constitutional rules influence the political process and ultimately outcomes such as the level, composition, and cyclicity of public spending, although the robustness of these findings has been challenged by Acemoglu (2005). In such cases, the usefulness of rules-based indicators as measures of governance depends crucially on how strong are the empirical links between such rules and the ultimate outcomes of interest.

Perhaps more common is the less extreme case in which rules-based indicators of governance do have normative content on their own, but the relative importance of different rules for outcomes of interest is unclear. The Global Integrity Index for example provides information on the existence of dozens of rules, ranging from the legal right to freedom of speech, to the existence of an independent ombudsman, to the presence of legislation prohibiting the offering or acceptance of bribes. The Open Budget Index provides highly-detailed factual information on the budget processes, including the types

of information provided in budget documents, public access to budget documents, and the interaction between executive and legislative branches in the budget process. Many of these indicators arguably have normative value on their own: having public access to budget documents is desirable by itself; and having streamlined business registration procedures is better than the alternative.

This leads to two related difficulties in using rules-based indicators to design and monitor governance reforms. The first is that absent good information on the links between changes in specific rules or procedures and outcomes of interest, it is difficult to know which of these rules should be reformed, and particularly in what order of priority. Will establishing an anticorruption commission or passing legislation outlawing bribery have any impact on reducing corruption, and if so, which one would be more important? Or should instead more efforts be put into ensuring that existing laws and regulations are implemented as intended, or that there is greater transparency and access to information, or greater media freedom? And how soon should we expect to see the impacts of one or more of these interventions? Given that governments typically operate with limited political capital to implement reforms, these tradeoffs and lags are important.

The second difficulty when designing or monitoring reforms arises when aid donors, or governments themselves, set performance indicators for governance reforms. Performance indicators based on changing specific rules, such as the passage of a particular piece of legislation, or a reform in a specific budget procedure, can be very attractive because of their clarity -- it is straightforward to verify whether the specified policy action has been taken.⁸ Yet it is important to underscore that "actionable" indicators are not necessarily also "action-worthy" in the sense of having a significant impact on the outcomes of interest. Moreover, excessive emphasis on registering improvements on rules-based indicators of governance leads to risks of "teaching to the test", or worse, "reform illusion", where specific rules or procedures are changed in isolation with the sole purpose of showing progress on the specific indicators used by aid donors.

⁸ Indeed, this is reflected in the terminology of "actionable" governance indicators emphasized in the World Bank's Global Monitoring Report (World Bank, 2006).

The final drawback of rules-based measures refer to the major gaps between statutory laws "on the books" and their implementation in practice "on the ground". To take an extreme example, in all of the 41 countries covered by the 2006 Global Integrity Index, accepting a bribe is codified as illegal, and all but three countries have an anticorruption commission or similar agency (Brazil, Lebanon, and Liberia were the only exceptions). Yet there is enormous variation in perceptions-based measures of corruption across these countries: the same list of 41 countries covered by the Global Integrity Index includes the Democratic Republic of Congo which ranks 200th, and the United States which ranks 23rd, out of 207 countries on the WGI Control of Corruption Indicator for 2006.

Another example of the gap between rules and implementation that we have documented in more detail elsewhere compares the statutory ease of establishing a business with a survey-based measure of firms' perceptions of the ease of starting a business, across a large sample of countries.⁹ In industrialized countries, where often *de jure* rules are implemented as intended by law, unsurprisingly we found that these two measures corresponded quite closely. In contrast, in developing countries where too often there are gaps between *de jure* rules and their *de facto* implementation, we found the correlation between the two to be very weak; in such countries *de jure* codification of the rules and regulations required to start a business is not a good predictor of the actual constraints as reported by firms. Unsurprisingly, much of the difference between the *de jure* and *de facto* measures of the ease of starting a business in developing countries could be statistically explained by *de facto* measures of corruption, which subverts the fair application of rules on the books.

These three drawbacks, namely an inevitable role of judgment even in "objective" indicators; the complexity and lack of knowledge regarding the links from rules to outcomes of interest; and the gap between rules "on the books" and their implementation "on the ground", suggest that although rules-based governance indicators provide valuable information, on their own they are insufficient for the purposes of measuring governance. Rules-based measures need to be complemented by and used in conjunction with outcome-based indicators of governance. We turn to such indicators, and their particular strengths and weaknesses, next.

⁹ Kaufmann, Kraay, and Mastruzzi (2006).

Outcome-Based Governance Indicators

The right-hand panel of Table 1 lists a selection of indicators that measure governance outcomes. As we noted, the majority of existing governance indicators fall in this category. Moreover, several of the sources of rules-based indicators of governance also provide outcome-based measures. The Global Integrity Index is a clear example in this respect, as it pairs up indicators of the existence of various rules and procedures with indicators of their effectiveness in practice. It is not the only one, however. The Database of Political Institutions for example not only measures such constitutional rules as the presence of a parliamentary system, but also outcomes of the electoral process such as the extent to which one party controls different branches of government, or the fraction of votes received by the president. Similarly, the Polity-IV database records a number of outcomes, including for example the effective constraints on the power of the executive.

The remaining outcome indicators range from the highly specific to the quite general. The Open Budget Index is an example of the former, reporting data on over 100 different indicators of the budget process across countries, ranging from whether budget documentation contains details of assumptions underlying macroeconomic forecasts, to the documentation of budget outcomes relative to budget plans. Other somewhat less specific sources include the Public Expenditure and Financial Accountability Indicators constructed by aid donors with inputs of recipient countries, and several large cross-country surveys of firms including the Investment Climate Assessments of the World Bank, the Executive Opinion Survey of the World Economic Forum, and the World Competitiveness Yearbook of the Institute for Management Development, which ask firms fairly detailed questions about their various interactions with the state.

Examples of more general assessments of broad areas of governance include ratings provided by several commercial sources including Political Risk Services (PRS), the Economist Intelligence Unit, and Global Insight-DRI. PRS for example provides ratings in 10 areas that can be identified with governance, such as "democratic

accountability", "government stability", "law and order", and "corruption". Other examples include large cross-country surveys of individuals such as the Afro- and Latino-Barometer surveys or the Gallup World Poll, which ask quite general questions such as: "is corruption widespread throughout the government in this country?".

The main advantage of such outcome-based indicators is that they capture very directly the views of relevant stakeholders, who take actions based on these views. Governments, analysts, researchers, opinion- and decision-makers should, and very often do, care about public views on the prevalence of corruption, the fairness of elections, the quality of service delivery, and many other governance outcomes. In other words, outcome-based governance indicators, as distinct from indicators of specific rules that we have discussed above, provide direct information on the *de facto* outcome of how the *de jure* rules are actually implemented: the distinction between rules "on the books" and practice "on the ground".

But against this major strength there are also some significant limitations. The first we have already discussed at length above. Outcome-based indicators of governance, and particularly where they are general ones, can be difficult to link back to specific policy interventions that might influence these governance outcomes. This is the mirror image of the problem we discussed above: rules-based indicators of governance can also be difficult to relate to outcomes of interest. A related difficulty is that outcome-based governance indicators may be too close to ultimate development outcomes of interest, and so become less useful as a tool for research and analysis. To take an extreme example, the recently-released Ibrahim Index of African Governance includes a number of ultimate development outcomes such as per capita GDP, growth of GDP, inflation, infant mortality, and inequality. While such development outcomes are surely worth monitoring, including them in an index of governance risks making the links from governance to development tautological.

Another difficulty has to do with interpreting the units in which outcomes are measured. We have noted that rules-based indicators have the virtue of clarity -- either a particular rule exists or it does not. Outcome-based indicators by contrast are often measured on somewhat arbitrary scales. For example, a survey question might ask respondents to rate the quality of public services on a 5-point scale, with the distinction

between different scores on this scale at times left rather unclear and up to the respondent.¹⁰ In contrast, the usefulness of outcome-based indicators is greatly enhanced by the extent to which the criteria for differing scores are clearly documented. The World Bank's CPIA and the Freedom House indicators are good examples of outcome-based indicators based on expert assessments that provide a fairly specific documentation of the criteria used to assign specific scores on the indicators that they compile. And in the case of surveys, questions can be designed in ways that ensure that responses are easier to interpret: rather than asking respondents whether they think "corruption is widespread", one can also simply ask whether they have been solicited for a bribe in the past month.

We conclude this section contrasting rules and outcomes-based measures of governance with an example to illustrate some of the main advantages and disadvantages of the two types of measures. Figure 1 compares alternative indicators of democratic accountability, a key dimension of governance. On the horizontal axis we have a very broad outcome indicator, taken from the 2005 Voice of the People survey, a large cross-country household survey. It asks households to answer whether they think elections in their country are free and fair. On the vertical axis, the series in circles at the top is a rules-based indicator of the quality of electoral institutions, taken from Global Integrity. It consists of a factual assessment of the existence of a number of specific institutions related to elections, such as the existence of a legal right to universal suffrage, and the existence of an election monitoring agency.¹¹ A first lesson from this graph is that in some cases, rules-based measures of governance show remarkable little variation across countries, with all countries receiving scores close to 100, indicating perfect scores on the "de jure" basis of this important aspect of governance. For example, a legal right to vote exists in every country surveyed by Global Integrity as of 2005, and a statutorily-independent election monitoring agency exists in all but three

¹⁰ See King and Wand (2007) for a description of how this problem can be mitigated by the use of "anchoring vignettes" that seek to provide a common frame of reference to respondents to aid in the interpretation of the response scale. The basic idea is to provide an understandable anecdote or vignette describing the situation faced by a hypothetical respondent to the survey, for example "Miguel frequently finds that his applications to renew a business license are rejected or delayed unless they are accompanied by an additional payment of 1000 pesos beyond the stated license fee". Respondents are then asked to assess how big an obstacle corruption is for Miguel's business, using a 10-point scale. Since all respondents use the scale to assess the same situation, this can be used to "anchor" their responses to questions referring to their own situation.

¹¹ Measured as the average of 14 "in law" components of the Elections indicator of Global Integrity. The other series on the graph is an average of the 20 "in practice" components of the same indicator.

(Lebanon, Montenegro, and Mozambique). Second, a striking feature of the graph is that the links between this specific objective indicator of rules and the broad outcome of interest, citizen satisfaction with elections, is at best very weak indeed, with a correlation between the two measures that is in fact slightly negative.

Third, the graph also illustrates how outcome-based indicators explicitly focusing on the *de facto* implementation of rules can be useful. As we have noted, a noteworthy feature of Global Integrity is its pairing of indicators of specific rules with assessments of their functioning in practice. The second series on the vertical axis (in squares, with countries labeled) reflects the assessment of Global Integrity's expert respondents as to the *de facto* functioning of electoral institutions. This series is much more strongly correlated with the broad outcome measure of interest taken from the Voice of the People survey, at 0.46. Yet at the same time, this correlation is far from perfect, and this in turn reminds us of the importance of relying on a variety of different indicators, pairing both expert assessments as well as survey-based indicators of "de facto" outcomes..

4. Whose Views Should We Rely On?

In this section we discuss alternative types of respondents on whose views governance indicators are based. The primary distinction here is between governance indicators based on the views of experts, and indicators capturing the views of survey respondents of various types. There are many examples of expert assessments listed in Table 1. We have already noted how rules-based indicators of governance like Doing Business rely on the views of one or a few legal experts per country, typically located in the capital city, to interpret the regulatory framework across countries. A large variety of governance assessments are produced by experts on behalf of commercial risk rating agencies and non-governmental organizations. The Global Integrity Index and the Open Budget Index for example rely on a locally-recruited expert in each country to complete their detailed questionnaires about governance, subject to peer review. Commercial organizations like the Economist Intelligence Unit rely on a network of their local correspondents in a large set of countries to provide information underlying the ratings that they produce. Other advocacy organizations like Amnesty International, Freedom House, and Reporters Without Borders also rely on networks of respondents for the information underlying their assessments. Governments and multilateral organizations

are also major producers of expert assessments. Some of the most notable include the Country Policy and Institutional Assessments produced by the World Bank, by the African Development Bank, and also by the Asian Development Bank. Each one of these assessments is based on the responses of their country economists to a detailed questionnaire, which are then reviewed for consistency and comparability across countries. Other examples include the Public Expenditure and Financial Accountability (PEFA) indicators mentioned above.

We also identify several large cross-country surveys of firms and individuals that contain questions relating to governance. These include the Investment Climate Assessment and the Business Environment and Enterprise Performance Surveys of the World Bank, the Executive Opinion Survey of the World Economic Forum, the World Competitiveness Yearbook, Voice of the People, and the Gallup World Poll.

Expert Assessments

Expert assessments have several major advantages which account for their preponderance among various types of governance indicators. One is simply cost: it is for example much less expensive to ask a selection of country economists at the World Bank to provide responses to a questionnaire on governance as part of the CPIA process than it is to carry out representative surveys of firms or households in a hundred or more countries. A second straightforward advantage is that expert assessments can more readily be tailored towards cross-country comparability: many of the organizations listed in Table 1 have fairly elaborate benchmarking systems to ensure that scores are comparable across countries. And finally, for certain aspects of governance, experts simply are the natural respondent for the type of information being sought. Consider for example the Open Budget Index's detailed questionnaire regarding national budget processes, the particulars of which are not the sort of common knowledge that survey data can easily collect.

Expert assessments nevertheless have several important limitations. A basic one is that, just as is the case among survey respondents, different experts may well have different views about similar aspects of governance. While this is perhaps not very surprising, it suggests that users of governance indicators should be cautious about

relying overly on any one set of expert assessments. We can get a particularly clean illustration of potential differences of opinion between expert assessments by comparing the CPIA ratings of the World Bank and the African Development Bank. These two institutions have in recent years harmonized their procedures for constructing CPIA ratings. Essentially, an identical questionnaire covering 16 dimensions of policy and institutional performance is completed by two very similar sets of expert respondents, namely country economists with in-depth experience working on behalf of these two organizations in the countries they are assessing.

Despite the homogeneity of the respondents and the very similar rating criteria, there are non-trivial differences between both organizations in the resulting assessments on the 16 components of the CPIA. Consider for example CPIA question 16 on "Transparency, Accountability, and Corruption in the Public Sector". The data for 2005 from both organizations are publicly available for a set of 38 low-income countries in Africa.¹² As reported in Table 2, the correlation between these two virtually identical expert assessments, while unsurprisingly positive, at 0.67 is nevertheless quite far from perfect. In the next section of the paper we discuss in more detail how we can interpret such differences of opinion as measurement error in each of the assessments, and how to quantify the extent of this measurement error. For now, however, we do note a very simple practical implication: when even very similar experts can provide significantly different assessments, it seems prudent to base assessments of governance for policy purposes on the views of a variety of different expert assessments.

Another critique often leveled against expert assessments of governance is just the opposite of the one we have discussed: that the country ratings assigned by different groups of experts are too highly correlated. The point here is a simple one. Suppose that one set of experts "does their homework" and comes up with an assessment of governance for a set of countries based on their own independent research, but a second set of experts simply reproduces the assessments of the first. In this case, the high correlation of two expert assessments cannot be interpreted as evidence of their accuracy. Rather, it would reflect the fact that the two sources make correlated errors in measuring governance. A priori, this should be a question of

¹² Starting with the 2005 data, both the African Development Bank and the World Bank have made public their CPIA scores. The AfDB does so for all borrowing countries while the World Bank does so only for countries eligible for its most concessional lending.

considerable concern.¹³ In this extreme example, we would in reality only have one data source, not two, and inferences about governance based on the two data sources would be no more informative than inferences based on just one of them.

This example is of course contrived because it makes the implausible assumption that the two data sources make perfectly correlated measurement errors when they assess governance across countries. However, even if the errors made by the two data sources are highly, but not perfectly, correlated, there will be benefits to relying on both of the data sources. The important empirical question is whether this hypothetical correlation of errors across sources is large or not. Empirically identifying correlations in errors across sources is difficult. Simply observing that two data sources provide assessments that are highly correlated is not enough, since the high correlation could reflect either (i) the fact that both sources are measuring governance accurately and so are highly correlated, or (ii) the fact that both sources are making correlated measurement errors in their assessments of countries.

In order to make progress we need to make identifying assumptions. In Kaufmann, Kraay and Mastruzzi (2006) we detail two sets of assumptions that allow us to disentangle potential sources of correlation in the errors. One assumption is that surveys of firms or individuals are less likely to make errors that are correlated with other data sources than, for example, the assessments of commercial risk rating agencies. If this is the case, however, we would expect that the assessments of commercial risk rating agencies be very highly correlated with each other, but less so with surveys. This turns out not to be the case. For example, the average correlation among our five major commercial risk rating agencies for corruption in 2002-2005 was 0.80. The correlation of each of these with a large cross-country survey of firms was actually slightly higher at 0.81, in contrast with what one would expect if the rating agencies had correlated errors. We do this exercise for components of all six of our aggregate governance indicators, and find at most quite modest evidence of error correlation. While this is unlikely to be the final word on this important question, we do think it is a useful step forward to

¹³ In fact, in our very first methodological paper on the aggregate governance indicators (Kaufmann, Kraay and Zoido-Lobaton 1999a) we devoted an entire section of the paper to this possibility, and showed how the estimated margins of error of our aggregate governance indicators would increase if we assumed that the error terms made by individual data sources were correlated with each other. Recently this critique has been raised again by Svensson (2005), Knack (2006) and Arndt and Oman (2006), although largely without the benefit of systematic evidence. Kaufmann, Kraay, and Mastruzzi (2007) provide a detailed response.

propose and implement tests of error correlation based on explicit identifying assumptions.

A third criticism of expert assessments is that they are subject to various biases. One argument is that many of these sources are biased towards the views of the business community, which may have very different views of what constitutes good governance than other types of respondents. In short, goes the critique, businesspeople like low taxes and less regulation, while the public good demands reasonable taxation and appropriate regulation. We do not think this critique is particularly compelling. If this is true, then the responses of commercial risk rating agencies who serve mostly business clients, or the views of firms themselves, to questions about governance should not be very correlated with ratings provided respondents who are more likely to sympathize with the common good, such as individuals, NGOs, or public sector organizations. Yet in most cases these correlations are in fact quite respectable. In Kaufmann, Kraay, and Mastruzzi (2007, Table 1) we document a strong correspondence between business-oriented sources of data on government effectiveness and other types of data sources. And in this paper, a glance at Table 2 suggests that cross-country surveys of firms and cross-country surveys of individuals, such as the World Economic Forum's Executive Opinion Survey and the Gallup World Poll result in similar rankings of countries according to views of corruption, with the two surveys correlated at 0.7 across countries.

Another potential source of bias in expert assessments, particularly those produced by NGOs, is that they are colored by the ideological orientation of the organization providing the ratings. In Kaufmann, Kraay, and Mastruzzi (2004) we devised a simple test for such political biases. We examined whether the difference between the assessments of think-tanks and firm surveys was systematically correlated with the political orientation of the government in power in the countries being rated. We found that this was generally not the case, casting doubt on this possible source of bias. Potentially a greater problem of bias is at the country respondent level. For example, in a particular country, the views of a pro-government and an anti-government "expert" might be very different, and this could affect both levels and trends over time in the scores for that country. This risk is perhaps greatest for sources that rely on locally-recruited experts, such as the Global Integrity Index. This is also much more difficult to

devise systematic statistical tests for, as the biases might affect individual country scores in one direction or another without introducing systematic biases into the source as a whole. Nevertheless, careful comparisons of many different data sources can often turn up anomalies in a single source that require more careful scrutiny.

Surveys of Firms and Individuals

We now turn to governance indicators derived from surveys of firms and individuals. Such indicators have the fundamental advantage that they elicit the views of the ultimate beneficiaries of good governance, citizens and firms in a country. Well-crafted survey-based governance indicators can capture the *de facto* reality on the ground facing firms and individuals, which as we have discussed above can be very different from the *de jure* rules on the books. The views of these stakeholders matter because they are likely to act on those views. If firms or individuals believe that the courts and the police are corrupt, they are unlikely to try to use their services (Hellman and Kaufmann (2004)) Individuals are less likely to vote, and to hold their elected leaders accountable, if they think that elections are not free and fair.

A further advantage of governance indicators based on surveys of domestic firms and individuals is their greater domestic political credibility. Governments can and do often dismiss external expert assessments of governance as uninformed pontification by outsiders. But it is much harder for governments to dismiss the views of their own citizens, or of firms operating in their country, when these point to failures of governance. Survey-based data on governance can therefore be particularly useful in galvanizing the politics of governance reforms. The experience of many countries implementing their own in-depth Governance and Anti-Corruption Diagnostics (assisted by the World Bank Institute and other agencies, and implemented with institutions in the requesting country), based on in-country surveys of enterprises, of users of services, and of public officials, supports this point: the reports on their views and experiences about many governance dimensions provided by thousands of stakeholders in the country provide a powerful input for action to reformist policy-makers and civil society groups.

Set against these important advantages of surveys there are again a number of disadvantages. First, we have the usual array of potential problems with any type of

survey data, ranging from issues of sampling design to issues of non-response bias. We note however the distinction with expert assessments, which by definition are based on the views of a very small number of respondents and so are less likely to be representative of the population of firms or households.¹⁴ While these generic issues are important for all surveys, we focus here on difficulties specific to measuring governance using survey data.

One disadvantage is that some survey questions on governance can be especially vague and open to interpretation, although as we discuss below, many have improved. An interesting example of this comes from innovative recent work by Razafindrakoto and Roubaud (2006). They use specially-designed surveys in eight African countries to contrast corruption perceptions based on household surveys with those based on expert assessments. The unique feature of this exercise is that the experts were asked to predict the country-level average responses from the household survey. In this sample of eight countries it turns out that the experts' ratings were essentially uncorrelated with the household survey responses. The authors conclude that the household surveys capture the "objective reality" of petty corruption and that the experts are just plain wrong. While this is a creative effort, we disagree with their interpretation that there is measurement error only in the expert assessment and not in the household survey. Households were asked whether they had been a "victim of corruption". There are a variety of reasons why households might think they were victimized by corruption when in fact it was not actually present. For example, a patient waiting in the queue to see a state-provided doctor might think (incorrectly) that people at the head of the queue had bribed someone to get there. Conversely households might well have paid a bribe, received the associated benefit, and found themselves quite satisfied and not at all "victimized" by the transaction. Our rather more modest interpretation of their finding is that there likely is measurement error in both the household survey, and in the matching expert assessments. And moreover, as we

¹⁴ This is not to say that all of the surveys used to measure governance are necessarily representative in any strict sense of the term. In fact, one general critique we note is that several of the large cross-country surveys of firms that provide data on governance are not very clear about their sample frame and sampling methodology. The Executive Opinion Survey of the World Economic Forum for example states that they seek to ensure that their sample of respondents is representative of the sectoral and size distribution of firms (World Economic Forum, 2006, p. 127). But at the same time they report that they "carefully select companies whose size and scope of activities guarantee that their executives benefit from international exposure" (World Economic Forum, 2006, p. 133). It is not clear from their documentation how these two conflicting objectives are reconciled.

discuss below, we find that in many other cases expert assessments and household survey responses do in fact correlate quite well across much larger samples of countries.

We note also that well-designed survey questions regarding corruption have become increasingly specific. For example, in some years questions in the Executive Opinion Survey of the World Economic Forum have asked firms to specifically report the fraction of contract value solicited in bribes on public procurement contracts. Greater attention is also being paid to techniques that enable respondents to report more truthfully to sensitive questions. For example, questions about corruption put to firms are often prefaced by "in your experience, do firms like your own typically pay bribes for.....?". Innovative techniques such as randomized response methods are used to protect the confidentiality of individual responses by allowing respondents to "camouflage" their response to sensitive questions by generating some of their responses at random based on the outcome of a coin toss, although they have not yet been widely used in large cross-country surveys.¹⁵ A related concern has to do with surveys of firms or individuals carried out in authoritarian countries where respondents might legitimately be fearful of responding truthfully to any question that might be interpreted as critical of the government.

Another potential difficulty in cross-country surveys of firms and individuals are cultural biases. It is often argued that respondents in different countries might have different norms as to what does or does not constitute corruption, and so their responses are not comparable across countries. Presumably however these cultural biases should not be present in cross-country expert assessments that are deliberately designed to be comparable across countries. And in many cases it turns out that surveys and expert assessments tend to produce very similar cross-country rankings. In Table 6 of Kaufmann, Kraay and Mastruzzi (2006b) we document sizeable correlations between expert assessments and the World Economic Forum's Executive Opinion Survey, for six different dimensions of governance. And a glance at Table 2 provides similar examples as well: for example the correlation across countries between the assessments of WMO, a commercial rating agency, and the Executive Opinion Survey, regarding

¹⁵ See for example Azfar and Murrell (2006) for an assessment of the extent to which randomized response methods succeed in correcting for respondent reticence, and an innovative approach to using this methodology to weed out less-than-candid respondents.

corruption is 0.88. While culture undoubtedly matters for the interpretation of survey responses across countries, we do not think that this is a first-order difficulty with cross-country comparability in survey-based data on governance.¹⁶

In short, as we saw when comparing measures of rules and measures of outcomes, in the case of expert assessments versus survey respondents, both types of data have their own unique strengths and weaknesses. Since neither type of respondent is clearly superior for all purposes, we think it important to continue to rely on a diversity of data sources in both dimensions of our taxonomy of governance indicators.

5. Aggregate or Individual Indicators?

Our discussion so far has focused on the strengths and weaknesses of alternative types of individual governance indicators. In this part of the paper we turn to the question of whether and when it makes sense to combine various such individual indicators of governance into aggregate or composite indicators combining information from multiple sources. In Table 1 we provide three examples of such aggregate indicators, the Worldwide Governance Indicators (WGI) that we have produced in other work, the well-known Corruption Perceptions Index (CPI) of Transparency International, and the very recently-released Ibrahim Index of African Governance. The WGI consist of six aggregate indicators of governance covering over 200 countries, combining cross-country data on governance provided by 30 different organizations. The CPI measures only corruption, using a smaller set of data drawn from nine different organizations. The WGI Control of Corruption indicator uses these nine data sources used by the CPI, as well as 13 others not used in the CPI. The Ibrahim Index is an extremely broad collection of a variety of types of indicators, including a number of subjective indicators such as those used in the WGI, and the CPI itself; as well as a number of very broad development outcomes, including per capita income, growth, inequality, and poverty.

¹⁶ Another way to assess the importance of such biases is to contrast perceptions-based measures of governance with more objective proxies. In general this is difficult because purely objective proxies are often hard to come by. One interesting recent example can be found in Fisman and Wei (2007) who study the discrepancy between recorded imports of objects of art into the United States, and the exports reported by partner countries, interpreting the discrepancy as evidence of art smuggling. They find that this purely objective proxy for illegal activity is highly correlated with the WGI measure of corruption. However, the correlation is also far from perfect, and as we discuss in the next section this implies non-trivial margins of error in both measures. It is also interesting to note that this is an objective measure of a governance *outcome* (art smuggling), in contrast with most of the so called 'objective' measures we have discussed that focus on *rules* regarding governance.

This makes the Ibrahim Index by far the broadest indicator we survey, but also makes it difficult to think of as a pure governance indicator because it also contains many broad development outcomes as well.

Measurement Error is Everywhere

A major theme in our discussion up to this point is that all governance indicators have limitations which make them noisy or imperfect proxies for the concepts they are intended to measure. The presence of measurement error in all governance indicators that this implies is central to the rationale for constructing aggregate indicators, and so we begin by discussing it in some detail. We think it is useful to distinguish between two broad types of measurement error that affect all types of governance indicators:

- First, any specific governance indicator will itself have measurement error relative to the particular concept it seeks to measure, due to intrinsic measurement challenges. For example, a survey question about corruption will have the usual sampling error associated with it. Similarly, we have already discussed how efforts to objectively document the specifics of the institutional environment or regulatory regime will face challenges in coming up with a factually accurate description of the relevant laws and regulations in each setting. Or for instance measures of the composition and volatility of public spending, which are sometimes interpreted as indicators of undesirable policy instability, are subject to all of the usual difficulties in measuring public spending consistently across countries and over time. And finally we have also noted how there can simply be differences of opinion between respondents -- for example different groups of experts might come up with rather different assessments of the same phenomenon in a particular country. These divergences of opinion can also usefully be interpreted as measurement error.
- Second, to the extent that we are interested in broad concepts of governance, any specific indicator is almost by definition an imperfect measure of the broader concepts to which it pertains, no matter how accurate or reliable that specific indicator is. A specific assessment of corruption in public procurement would not be fully informative about overall corruption in the public sphere, even if it were

fully accurate about this specific type of corruption. Information about the statutory requirements for business entry regulation need not reflect the actual practice of how these requirements are implemented on the ground, nor are they informative about regulatory burdens in other areas. Information about freedom of the press is only one of many factors contributing to the accountability of governments to their citizens. Notwithstanding some clear advantages that specificity of an indicator may have for some purposes, one should be careful not to interpret them as sufficient statistics for broader notions of governance.

How important is this measurement error quantitatively? Unfortunately, the vast majority of existing governance indicators do not explicitly acknowledge the extent of measurement error present in them. One of the few exceptions to this is the Worldwide Governance Indicators (WGI) project that we discuss further below. Fortunately, however, some simple calculations can shed some light on the likely magnitude of measurement error in individual governance indicators as well. The key to doing so is to identify pairs of indicators that arguably are measuring similar concepts, up to an unavoidable measurement error component. For example we have discussed above the correlation between the World Bank and African Development Bank's CPIA assessments of transparency and corruption. A useful way to interpret this imperfect correlation is that both of these sources are measuring the same concept of transparency, accountability, and corruption, but they do so with a degree of measurement error. Intuitively, the less measurement error there is in these two sources, the more correlated they should be. Thus we can interpret the correlation between them as telling us something about the degree of measurement error that is present.

More formally, think of the observed scores from two organizations, y_1 and y_2 , as a combination of a signal about unobserved governance, g , and source-specific noise, ε_1 and ε_2 , i.e. $y_1 = g + \varepsilon_1$ and $y_2 = g + \varepsilon_2$. Suppose we assume that the variance of measurement error in the assessments of the two organizations is the same, and without loss of generality assume that the variance of governance is one.¹⁷ Then some simple

¹⁷ The assumption of a common error variance is necessary in this simple example with two indicators in order to achieve identification. In this example there is just one sample correlation in the data that can be used to infer the variance of measurement error: we thus can identify just one measurement error variance.

arithmetic tells us that the standard deviation of measurement error is

$SD(\varepsilon) = \sqrt{(1-\rho)/\rho}$ where ρ is the correlation between these two particular expert assessments.¹⁸ We report this standard deviation in Table 3 for several pairs of indicators that we have discussed in the paper, and it ranges from 0.70 to 1.53 across our examples. By way of comparison, the standard errors associated with aggregate indicators such as the WGI are much smaller, reflecting the benefits of aggregation in reducing noise in the individual indicators. For example the standard error for the estimate of Control of Corruption for a typical country in 2006 on the WGI is just 0.17, or less than a quarter of the standard error of the most precise pair of individual indicators in this example.

To appreciate the magnitude of this measurement error, it is useful to go one step further and calculate the width of a 90% confidence interval for governance based on any one of these individual indicators, based on the additional assumption that governance and the error term are jointly normally distributed. This width of the confidence interval is $2 \times 1.64 \times SD(g | y) = 3.28 \times \sqrt{1-\rho}$, and is reported in the last column of Table 3. Since our assumptions imply that 95 percent of countries would have governance levels between -2 and 2, these figures imply that a 90 percent confidence interval for governance for any individual country would span between half and two-thirds of the entire most-likely range of governance outcomes!

Why Aggregate Indicators?

We have emphasized how measurement error is present in all indicators of governance. Aggregate indicators of governance can be a useful way of combining, organizing, and summarizing the information from alternative sources, and thus reducing the influence of measurement error in any individual indicator. This is no more than a simple intuition that averaging across different proxies for governance will provide a

In more general applications of the unobserved components model, such as the Worldwide Governance Indicators, this restriction is not required as there are three or more data sources.

¹⁸ For details on this calculation see Kaufmann, Kraay, and Mastruzzi (2004, 2006). Gelb, Ngo and Ye (2004) implement a similar calculation comparing the African Development Bank and World Bank CPIA scores. Their conclusion that the CPIA ratings are quite precise is largely driven by the fact that they focus on the aggregate CPIA scores (across all macro, structural, social and public sector dimensions), which are very highly correlated between the two institutions. In the case of the CPIA, here we focus on one of 16 specific questions, and at this level of disaggregation the correlation between the two sets of ratings is considerably lower.

more informative measure of governance than any individual indicator. A further significant benefit of aggregation, however, is that it allows for the construction of explicit margins of error for both the aggregate indicator itself as well as its component individual indicators.

The Worldwide Governance Indicators (WGI) that we have developed over the past decade illustrate how these margins of error can be calculated (refer also to Box 1). In particular, the statistical methodology underpinning the Worldwide Governance Indicators, the unobserved components model, explicitly assumes that the true level of governance is unobservable, and that the observed empirical indicators of governance provide noisy or imperfect signals of the fundamentally unobservable concept of governance. This formalizes what we have been discussing throughout this survey -- that all available indicators are imperfect proxies for governance. The estimates of governance that come out of this model are simply the conditional expectation of governance in each country, conditioning on the observed data for each country. Moreover, the unobserved components model allows us to summarize our uncertainty about these estimates for each country with the standard deviation of unobserved governance, again conditional on the observed data. These can be used to construct confidence intervals for governance estimates, which we often refer to informally as "margins of error". Intuitively, these margins of error for our estimates of governance are smaller the more data sources are available for a given country. We can also estimate the variance of the error term in each individual underlying governance indicator using this methodology, following a calculation that generalizes the simple one we discussed above.

From the standpoint of users, these margins of error associated with estimates of governance are non-trivial, as illustrated in Figure 2. The graph reports selected countries' scores on the WGI Control of Corruption indicator, for 2006. The height of the bars denotes the estimates of corruption for each country, and the thin vertical lines on each bar denote 90 percent confidence intervals. For many pairs of countries with similar scores, these confidence intervals overlap, indicating that the small differences between them are unlikely to be statistically, or practically significant. However, we do also note that there are many possible pair-wise comparisons between countries that do result in significant differences. Roughly two-thirds of the possible pair-wise

comparisons of corruption across countries using this indicator result in differences that are significant at the 90 percent confidence level, and nearly three-quarters of comparisons are significant at a less-demanding 75 percent confidence level. Clearly, far fewer pair-wise comparisons would be significant if they were based on any single individual indicator whose margins of error have not been reduced by averaging across alternative data sources. For example, if we take an individual data source with a typical standard error from the WGI Control of Corruption indicator, such as Global Insight-DRI, we find that only 16 percent of cross-country comparisons based on this one data source would be significant at the 90 percent confidence level.

As we have noted, however, the WGI are unusual among existing governance indicators in their transparent recognition of such margins of error. The vast majority of investment climate and governance indicators simply report country scores or ranks without any effort to quantify the measurement error that these rankings inevitably contain. This has tended to contribute to a sense of spurious precision among users of these indicators, and to an overemphasis of small differences between countries.

Of course aggregate indicators have their own shortcomings as well. Foremost among these is an inevitable loss of specificity that comes with aggregate indicators of governance. If we average one indicator of judicial corruption and another indicator of bureaucratic corruption, we arguably have a more informative indicator of overall corruption, but we do not have a more informative indicator of either of the two particular types of corruption. Similarly if we average an indicator of press freedoms with an indicator of electoral integrity we have a more informative indicator of overall democratic accountability but we do not have any more precise an indicator of either of these two specific concepts. For some purposes the broad aggregate indicators will be useful, while for other purposes the disaggregated underlying indicators will be more useful. However, we do not view this as a major shortcoming, since after all, virtually all aggregate governance indicators can readily be disaggregated into their constituent components, giving the user the freedom to choose the appropriate level of aggregation for the task at hand.¹⁹

¹⁹ In the case of the WGI, the full dataset of individual indicators underlying the aggregate indicators is available through an interactive website at www.govindicators.org.

A second concern with aggregate indicators is that their effectiveness at reducing measurement error depends crucially on the extent to which their underlying sources provide independent information on governance. We have already discussed in Section 2 the criticism that some types of expert assessments might make correlated errors in their governance rankings, although empirical evidence suggests that these error correlations are likely not very large in practice. However, it is important to keep in mind that aggregate indicators can only mitigate the component of measurement error that is truly independent across the different underlying indicators. This point is particularly relevant when contrasting "multiple-source" and "single-source" aggregate indicators. The Worldwide Governance Indicators are an example of the former, combining information from a large number of distinct data sources. In contrast, many of the data sources reported in Table 1 report aggregates of their own subcomponents. For example, there is an aggregate CPIA rating in conjunction with the 16 underlying components, and there are six aggregate Global Integrity indicators combining information from over 200 underlying individual indicators. The distinction here is that in the latter case, all of the underlying individual indicators for a given country are scored by the same respondent. As a result, any respondent-specific biases are likely to be reflected in all of the individual indicators, and so the gain in precision from relying on the aggregate indicators from these sources will not be as large as when aggregate indicators are based on multiple underlying sources.

In summary, we have argued that aggregate governance indicators can play a useful role in synthesizing and summarizing the large variety of existing individual governance indicators. In this sense, the use of aggregate indicators is one way to exploit the complementarities between the different types of indicators that we have discussed in the previous sections (rules vs. outcomes, surveys vs. experts). A further benefit of aggregation is the increase in precision with which these aggregate indicators measure broad, although unobservable, concepts of governance. At the same time however we recognize that for some purposes more specific indicators are useful, and thus it is important to be able to easily disaggregate aggregate indicators into their constituent components, as for example is the case with the WGI.

6. Moving Forward

In this paper we have taken stock of existing cross-country indicators of governance, using a simple framework based on two questions: "what do we measure?" and "whose views do we rely on?". We emphasized the distinction between rules-based and outcome indicators, as well as the distinction between drawing data from experts or respondent surveys of citizens or enterprises. We also discussed the rationale for aggregate indicators, noting that different levels of aggregation are appropriate for different purposes. A sobering perspective emerges from this review, arguing for circumspection: while most indicators have many virtues, all face distinct challenges as well. This points to the need to look at a variety of indicators and sources when monitoring or assessing governance across countries, within a country and over time. We conclude by offering a few principles that may be useful as this work, and the use of governance indicators in public sector policymaking and civil society monitoring, continues.

Avoid false dichotomies. Too often discussions of governance indicators overemphasize distinctions between alternative types of governance indicators, with insufficient regard for the strong complementarities between them. For example, artificially sharp distinctions are often drawn between so called "subjective" and "objective" indicators of governance. As we have discussed however, virtually all indicators of governance rely on the judgments or perceptions of respondents in one way or another, and so we suggest that this distinction is largely artificial. And in some cases even the terminology is used in misleading ways. For example, the very recently-released Ibrahim Index of African Governance touts itself as providing "objective" assessments of governance despite the fact that it is based on a number of purely subjective data sources including the Transparency International Corruption Perceptions Index and subjective ratings produced by the Heritage Foundation and the Economist Intelligence Unit.

Distinctions made between aggregated and disaggregated indicators often have an artificial element as well. First, some aggregate indicators transparently disclose each disaggregated source, enabling users to take advantage of the complementarities between both types and blurring the distinction between the two categories. Further, for

some purposes it is useful to combine information from many individual indicators into some kind of summary statistic, while for other purposes the disaggregated data are of primary interest. Even where disaggregated data are of primary interest, however, it is important to rely on a number of independent sources for validation, since the margins of error, and the likelihood of extreme outliers, is significantly higher for a disaggregated indicator.

An excessively narrow emphasis on "actionable" indicators detailing specific policy interventions immediately under the control of governments can divert attention from equally-important discussions of which of these indicators are "action-worthy" in the sense of having significant impacts on outcomes of interest. And the answer to which indicators are most "action-worthy" is rarely obvious a priori, and often context-specific. Focusing excessively on "actionable" indicators, while downplaying the scrutiny on outcome indicators, may result in undue emphasis on measures that may not translate into concrete progress on outcomes.

Use indicators appropriate for the task at hand. As with all tools, different types of governance indicators are suited for different purposes. In this survey we have emphasized governance indicators that can be used for regular cross-national comparisons. While many of these indicators have become increasingly specific over time, often they remain blunt tools for monitoring governance and studying the causes and consequences of good governance at the country level. For these purposes a wide variety of innovative tools and methods of analysis have been deployed at the country level in many countries worldwide, and reviewing these is well beyond the scope of this survey. Examples among these in-country tools include the World Bank's Investment Climate Assessments (ICAs), the World Bank Institute's Governance and Anti-Corruption (GAC) diagnostics, the corruption surveys carried out by some chapters of Transparency International (TI), and the institutional scorecard carried out by the Public Affairs Center in Bangalore, India. And disaggregating further to the level of individual projects, many project-specific interventions and diagnostics are possible to measure governance carefully at this level.²⁰

²⁰ One of the best-known and best-executed recent studies of this type is a study of corruption in a local road-building project by Olken (2007).

Public and professional scrutiny is essential for the credibility of governance indicators. Virtually all of the governance indicators listed in Table 1 are publicly-available, either commercially or at no cost to users. This transparent feature is central to their credibility as tools to monitor governance. Open availability permits broad scrutiny and public debate about the content and methodology of indicators and their implications for individual countries. Many indicators are also produced by non-government actors, making it more likely that they are immune from either the perception or the reality of self-interested manipulation on the part of governments being assessed. Scholarly peer review can also help to strengthen the quality and credibility of governance indicators. For example, a number papers describing the methodology of the Doing Business indicators, the Database of Political Institutions, and the Worldwide Governance Indicators have all appeared in peer-reviewed professional journals. Transparency with respect to details of methodology and its limitation is also essential for credible use of governance indicators. It is important that users of governance indicators understand fully the characteristics of the indicators they are using, including any methodological changes over time as well as time lags between the collection of data and publication.

It is thus of concern that some proposed and existing indicators of governance are as yet insufficiently open to public scrutiny. While the recent disclosure of the World Bank's CPIA ratings for low-income countries is a very useful step, these indicators are now being disclosed for only about half of the roughly 130 countries for which they are prepared each year, and not at all for the historical data prior to 2005. Similarly, historical data on the CPIA assessments of the African Development Bank and Asian Development Bank have not been disclosed publicly. This is unfortunate given that the decision to only selectively disclose recent CPIA data, and not to disclose historical CPIA data, is made by the executive boards of these organizations and so reflects the desire of the very governments these ratings are supposed to assess. Regarding transparency, it is also of concern that although the Public Expenditure and Financial Accountability initiative, which has been ongoing for seven years, has so far resulted in indicators and reports being constructed for just 42 countries as of March 2007, for only one period per country, and only nine of these are publicly available. Moreover, since these reports are prepared in collaboration with the governments in question, their credibility would not be the same as those associated with purely third-party indicators.

Similar concerns affect recent OECD-led efforts to construct indicators of public procurement practices.

Transparently acknowledge margins of error of all governance indicators. All governance indicators have measurement error and so should be thought of as imperfect proxies for the fundamentals of good governance that we seek to measure and improve across countries. This is not just an abstract statistical point, but rather one of fundamental importance for all users of governance indicators. Wherever possible such margins of error should be explicitly acknowledged, as for example they are in the Worldwide Governance Indicators project. And these margins of error should be taken seriously when using these indicators to monitor progress on governance. At times the lack of disclosure of margins of error is rationalized by suggesting that they would be largely missed by most readers. Yet our experience with the Worldwide Governance Indicators suggests that this is not the case, with many users recognizing and benefitting from this additional degree of transparency about data limitations.

Exploit the wealth of currently-available indicators, recognizing that further progress in developing new governance indicators is likely to be incremental. As we have seen in this survey, a very large number of different governance indicators already exist. Considerable research has been done on the cross-national links between broad measures of governance and broad development outcomes. But much more work needs to be done to exploit the large body of disaggregated measures of governance already in existence. Linking disaggregated indicators to disaggregated outcomes, both across countries and over time, is likely to be an important and exciting area of research over the next several years. And it is likely to have important implications for policymakers.

At the same time, there is also scope for developing new and better indicators of governance to address some of the weaknesses of existing measures that we have flagged in this review. Work to improve such indicators will be important as indicators are increasingly used to monitor the success and failure of governance reform efforts. But given the many challenges of measuring governance, it is also important to recognize that progress in this area over the next several years is likely to be incremental rather than fundamental. In fact, in terms of potential payoff, alongside

efforts to develop new indicators there is also a case to improve upon existing indicators, particularly in increasing the periodicity of heretofore one-off efforts, and broadening their country coverage (covering industrialized and developing countries), as well as covering issues for which data is still scarce, such as money laundering.

References

- Acemoglu, Daron (2006). "Constitutions, Politics, and Economics: A Review Essay on Persson and Tabellini's *The Economic Effects of Constitutions*". *Journal of Economic Literature*. 63(4):1025-1048.
- Arndt, Christiane and Charles Oman (2006). "Uses and Abuses of Governance Indicators". OECD Development Center Study.
- Azfar, Omar and Peter Murrell (2006). "Identifying Reticent Respondents: Assessing the Quality of Survey Data on Corruption and Values". Manuscript, University of Maryland.
- Gelb, Alan, Brian Ngo, and Xiao Ye (2004). "Implementing Performance-Based Aid in Africa: The Country Policy and Institutional Assessment". World Bank Africa Region Working Paper No. 77.
- Fisman, Raymond and Shang-jin Wei (2007). "The Smuggling of Art and the Art of Smuggling: Uncovering Illicit Trade in Cultural Property and Antiques". Manuscript, Columbia University.
- Hellman, Joel and Daniel Kaufmann (2004). 'The Inequality of Influence', in Kornai, J. and S. Rose-Ackerman, eds., *Building a Trustworthy State in Post-Socialist Transition*, Palgrave MacMillan
- Kaufmann, Daniel, Aart Kraay and Pablo Zoido-Lobaton (1999a). "Aggregating Governance Indicators." World Bank Policy Research Working Paper No. 2195, Washington, D.C.
- Kaufmann, Daniel, Aart Kraay and Pablo Zoido-Lobaton (1999b). "Governance Matters." World Bank Policy Research Working Paper No. 2196, Washington, D.C.
- Kaufmann, Daniel, Aart Kraay and Massimo Mastruzzi (2006). "Governance Matters V: Governance Indicators for 1996-2005. World Bank Policy Research Department Working Paper No. 4012.
- Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi (2007a). "The Worldwide Governance Indicators Project: Answering the Critics". World Bank Policy Research Department Working Paper No. 4149.
- Kaufmann, Daniel, Aart Kraay and Massimo Mastruzzi (2006). "Governance Matters V: Aggregate and Individual Governance Indicators for 1996-2006. World Bank Policy Research Department Working Paper No. 4280.
- Kautilya (400 B.C.) *The Arthashastra*
- Knack, Steven (2006). "Measuring Corruption in Eastern Europe and Central Asia: A Critique of the Cross-Country Indicators". World Bank Policy Research Department Working Paper 3968.

- King, Gary and Jonathan Wand (2007). "Comparing Incomparable Survey Responses: Evaluating and Selecting Anchoring Vignettes". *Political Analysis*. 15(1):46-66.
- North, Douglass (2000). "Poverty in the Midst of Plenty". Hoover Institution Daily Report, October 2, 2000. www.hoover.org.
- Olken, Ben (2007). "Monitoring Corruption: Evidence from a Field Experiment in Indonesia". *Journal of Political Economy*. 115(2):200-249.
- Persson and Tabellini (2005). *The Economic Effects of Constitutions*. Cambridge, MIT Press.
- Razafindrakoto, Mireille and Francois Roubaud (2006). "Are International Databases on Corruption Reliable? A Comparison of Expert Opinion Surveys and Household Surveys in Sub-Saharan Africa". Manuscript, IRD/DIAL, http://www.dial.prd.fr/dial_publications/PDF/Doc_travail/2006-17.pdf.
- Svensson, Jakob (2005). "Eight Questions About Corruption". *Journal of Economic Perspectives*. 19(3):19-42.
- UNDP (2005). *Governance Indicators: A Users Guide*.
- World Economic Forum (2006). "The Global Competitiveness Report 2006-2007". New York: Palgrave Macmillan.
- World Bank (1992). "Governance and Development". Washington.
- World Bank (2002). "Building Institutions for Markets". Washington: Oxford University Press.
- World Bank (2006). "Global Monitoring Report 2006: Strengthening Mutual Accountability: Aid, Trade, and Governance".
- World Bank (2007). "Strengthening World Bank Group Engagement on Governance and Anticorruption". <http://www.worldbank.org/html/extdr/comments/governancefeedback/gacpaper-03212007.pdf>

Table 1 -- Taxonomy of Governance Indicators

Taxonomy of Existing Governance Indicators

Whose Opinion?	About What?			
	<u>Rules</u>		<u>Outcomes</u>	
	<i>Broad</i>	<i>Specific</i>	<i>Broad</i>	<i>Specific</i>
<u>Experts</u>				
<i>Lawyers</i>		DB		
<i>Commercial Risk Rating Agencies</i>			DRI, EIU, PRS	
<i>Non-Governmental Organizations</i>		GII	HER, RSF, CIR, FRH	GII, OBI
<i>Governments and Multilaterals</i>			CPIA	PEFA
<i>Academics</i>	DPI, PIV		DPI, PIV	
<u>Survey Respondents</u>				
<i>Firms</i>				ICA, GCS, WCY
<i>Individuals</i>			AFR, LBO, GWP	
<u>Aggregate Indicators Combining Respondents</u>			TI, WGI, MOI	

Legend

<i>Code</i>	<i>Name</i>	<i>Countries Covered</i>	<i>Frequency</i>	<i>Link</i>
AFR	Afrobarometer	18	Every 3 years	
CIR	Cingranelli-Richards Human Rights Dataset	192	Annual	www.humanrightsdata.com
CPIA	Country Policy and Institutional Assessment	136	Annual	www.worldbank.org
DB	Doing Business	175	Annual	www.doingbusiness.org
DPI	Database of Political Institutions	178	Annual	http://econ.worldbank.org
DRI	Global Insight DRI	117	Quarterly	www.globalinsight.com
EIU	Economist Intelligence Unit	120	Quarterly	www.eiu.com
FRH	Freedom House	192	Annual	www.freedomhouse.org
GCS	Global Competitiveness Survey	117	Annual	www.weforum.org
GII	Global Integrity Index	41	Every 3 years	www.globalintegrity.org
GWP	Gallup World Poll	131	Annual	www.gallupworldpoll.com
HER	Heritage Foundation	161	Annual	www.heritage.org
ICA	Investment Climate Surveys	94	Irregular	www.investmentclimate.org
LBO	Latinobarometro	17	Annual	www.latinobarometro.org
MOI	Ibrahim Index of African Governance	48	Every 3 years	www.moibrahimfoundation.org
OBI	Open Budget Index	59	Annual	www.openbudgetindex.org
PEFA	Public Expenditure and Fiscal Accountability	42	Irregular	www.pefa.org
PIV	Polity IV	161	Annual	www.cidcm.umd.edu/polity/
PRS	Political Risk Services	140	Monthly	www.prsgroup.com
RSF	Reporters Without Borders	165	Annual	www.rsf.org
WCY	World Competitiveness Yearbook	47	Annual	www.imd.ch

Table 2: Correlation Among Alternative Indicators of Corruption

	Expert Assessments			Surveys		
	World Bank CPIA	AfDB CPIA	Global Integrity	World Markets Online	WEF Executive Opinion Survey	Gallup World Poll
World Bank CPIA	1.00	0.67	0.30	0.56	0.25	0.13
African Development Bank CPIA		1.00	0.49	0.51	0.45	0.24
Global Integrity			1.00	0.34	0.29	0.11
World Markets Online				1.00	0.88	0.59
WEF Executive Opinion Survey					1.00	0.70
						1.00

Table 3: Measurement Error in Individual Governance Indicators

	<u>Correlation</u>	<u>Standard Deviation of Error</u>	<u>Length of 90% Confidence Interval</u>
<i>Transparency, Accountability and Corruption</i> World Bank CPIA-16 African Development Bank CPIA-16	0.67	0.70	1.88
<i>Business Entry Regulation</i> EOS Survey Doing Business	0.48	1.04	2.37
<i>Elections</i> Global Corruption Barometer Survey Global Integrity Elections Index	0.30	1.53	2.74

Figure 1: De Jure and De Facto Indicators of Elections

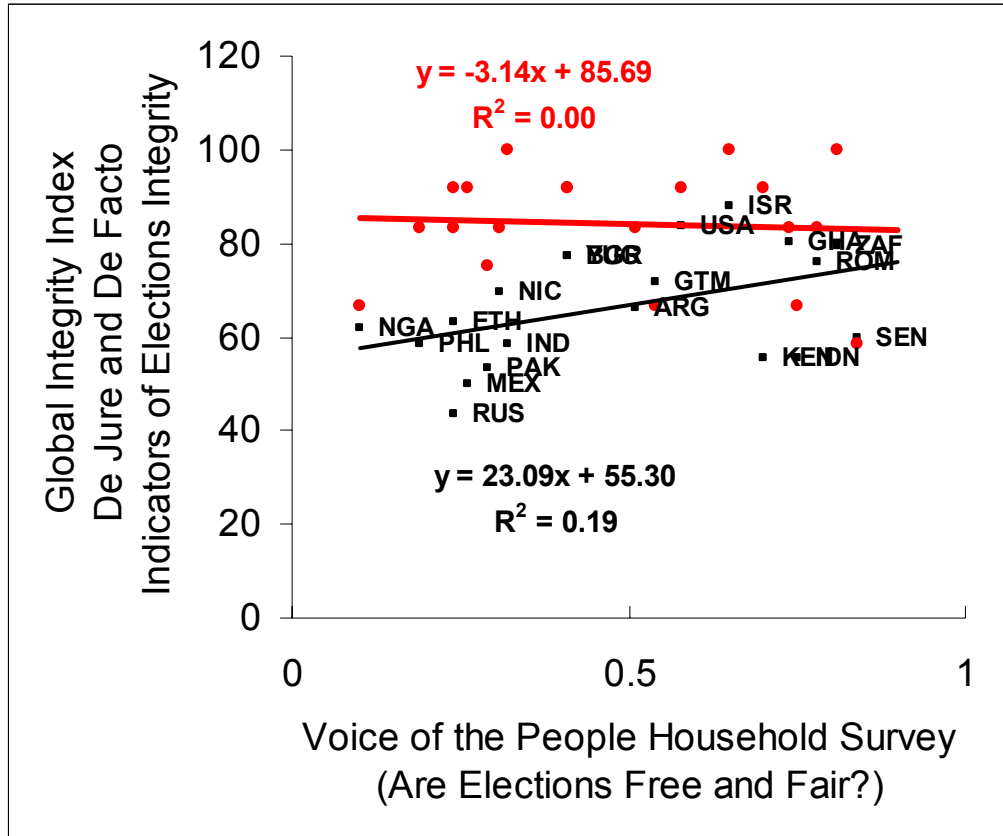
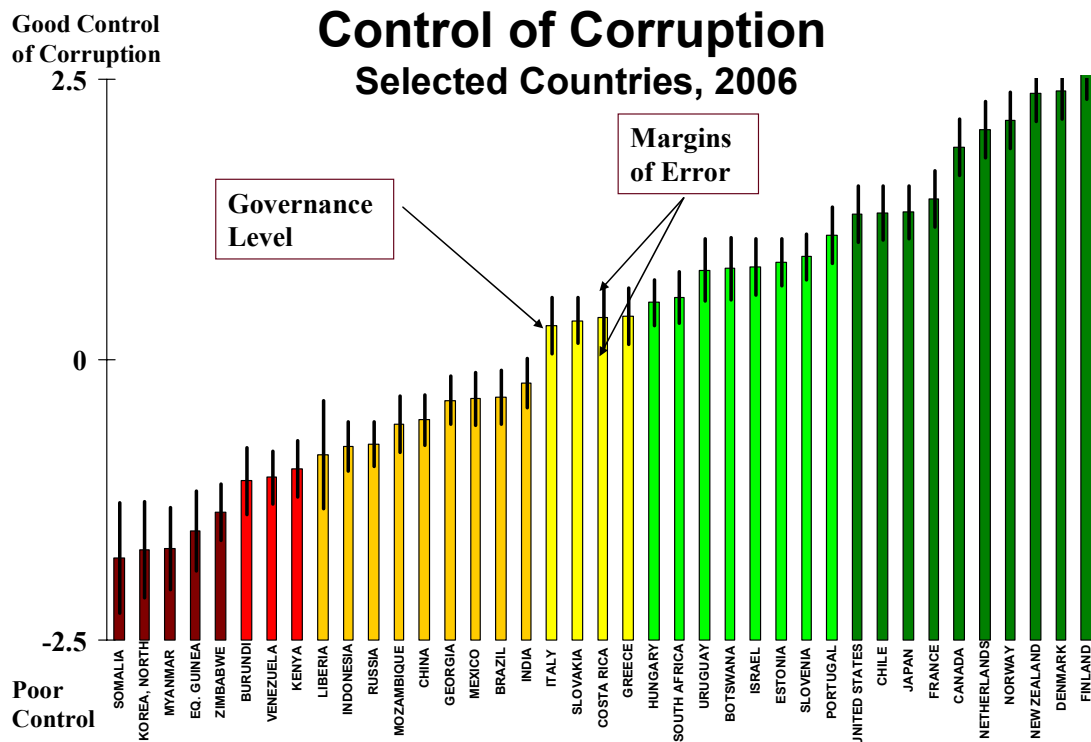


Figure 2: Margins of Error in Estimates of Governance



Source for data: 'Governance Matters VI: Governance Indicators for 1996-2006', by D. Kaufmann, A. Kraay and M. Mastruzzi, June 2007, www.govindicators.org. Colors are assigned according to the following criteria: Dark Red: country is in the bottom 10th percentile rank ('governance crisis'); Light Red: between 10th and 25th percentile rank; Orange: between 25th and 50th percentile rank; Yellow, between 50th and 75th; Light Green between 75th and 90th percentile rank; and Dark Green: between 90th and 100th percentile (exemplary governance). Estimates subject to margins of error.

Box 1: The Worldwide Governance Indicators: Critiques and Responses

The Worldwide Governance Indicators (WGI) are among the most widely-used cross-country governance indicators currently available (see Kaufmann, Kraay, and Mastruzzi (2007b) for a description of the latest update). The WGI report on six dimensions of governance for over 200 countries for the period 1996-2006, and are based on hundreds of underlying individual indicators drawn from 30 different organizations, relying on responses from tens of thousands of citizens, enterprise managers, and experts. The WGI have also attracted some specific written critiques. This box briefly summarizes the major critiques and our rebuttals. More details on these critiques and our responses can be found in Kaufmann, Kraay, and Mastruzzi (2007a).

Comparability over time and across countries. Several critics have raised concerns about the over-time and cross-country comparability of the WGI, noting that (i) the WGI use units that set the global average of governance to be the same in all periods; (ii) comparisons of pairs of countries, or single countries over time using the WGI will be based on different sets of underlying data sources; and (iii) there are substantial margins of error in the aggregate WGI. In response, we note that (i) we have documented for several years that there is no clear evidence of a clear trend in one direction or another in global averages of governance on any of our underlying individual data sources (the overall evidence pointing to general stagnation), so that the choice of a constant global average is no more than an innocuous choice of units; (ii) we have documented that changes in the set of underlying data sources on average contributes only minimally to changes over time in countries' scores on the aggregate WGI, and that the majority of cross-country comparisons using the aggregate WGI are based on a substantial number of common data sources; and (iii) we view the presence of explicit margins of error in the WGI as an important advantage of these indicators, serving as a useful antidote to superficial comparisons of country ranks or country performance over time that are often made with other governance indicators. Nevertheless, as we discuss in the main text in Section 5, a substantial fraction of cross-country and over-time comparisons using the WGI do result in statistically significant differences, suggesting that the WGI are in fact usefully informative.

Biases in Expert Assessments. Several critics have alleged biases of various sorts in the data sources underlying the WGI, including an excessive emphasis on business-friendly regulation on the part of some data providers; ideological biases such as a bias against left-wing governments on the part of some data providers; and "halo effects" whereby countries with good economic performance receive better-than-warranted governance scores. Providing empirical evidence in support of such biases is much more difficult, and in our view has not yet been done convincingly. In the main text in Section 4 we have reviewed some of our own empirical work which suggests that these biases, even where they a priori may be present, are quantitatively unimportant.

Correlated Perception Errors. Several critics have suggested that expert assessments make similar errors when assessing the same country, leading to correlations in the perception errors across various expert assessments. While this is plausible, there is little convincing empirical evidence in support of it, and in the main text in Section 4 we have reviewed some of our own empirical work which suggests that these biases are quantitatively unimportant. A related concern is that correlated perception errors will lead to an over-weighting of such sources in the aggregate WGI, since the WGI weights individual data sources by estimates of their precision, which in turn are based on the

observed intercorrelation among sources (see discussion in Section 5 of the main text). Given the at best modest evidence of correlated perceptions errors, this is unlikely to be quantitatively important. Further, we have also documented that the country rankings on the WGI are highly robust to alternative weighting schemes.

Definitional Issues. Some critics have taken issue with our definitions of governance, and thus the assignment of individual governance indicators to the six aggregate WGI. As we have discussed in the main text in Section 2, there is no sharp definitional consensus in the area of measuring governance, and so there cannot be "right" or "wrong" definitions, and corresponding measures, of governance. Nevertheless, most reasonable definitions of governance cover similar broad areas, and aggregate indicators capturing these broad areas are likely to be similar. Moreover, since virtually all of the individual indicators underlying the WGI are publicly available through the WGI website, researchers can easily construct alternative indicators corresponding to their preferred notions of governance.

Reliance on "Subjective" Data. Various critics have argued that the perceptions-based data on which the WGI are based do no more than reflect vague and generic perceptions rather than specific objective realities, and that "specific, objective, and actionable" measures of governance are needed to guide policymakers and to make progress in governance reforms. We have already discussed at length in this survey how virtually all governance indicators necessarily involve some element of subjectivity; how perceptions-based data on their own are extremely valuable in that they capture the views of relevant stakeholders who act on these views; and that the links from specific changes to policy rules are very difficult to link to changes in outcomes of interest, and so it is difficult to identify indicators that are "actionworthy" as opposed to merely being "actionable".